A quick overview of the Swendsen-Wang method and its application to Boltzmann machines

Simon Osindero

University of Toronto

Tea Talk: 14-July-2004
Several efficient Monte Carlo methods involve augmenting the original variables in a model with a set of auxiliary variables. Operating in this enlarged space can sometimes be much easier and more efficient than operating in the original space.

- One example of this type of method is the Hybrid Monte Carlo approach for systems of continuous variables.
- Swendsen-Wang methods (and their generalisations) are another example, and have been mostly applied to systems of discrete variables.
Motivation

- Sampling from Boltzmann machines (and similar MRF models) is difficult in the general case. Markov Chain Monte Carlo (MCMC) methods are often required.

- Unless we have particular structure (e.g., a bipartite graph as in Restricted Boltzmann Machines), the method commonly used is single variable updating using Gibbs sampling.
  - This may suffer from slow mixing, requiring many updates to produce significant changes in the global state.

- Enhanced methods that allow multiple variables to be updated simultaneously can potentially lead to faster mixing and greater efficiency.
Basic Idea: Introduce Auxiliary Variables

- Somewhat surprisingly, we can sometimes make a sampling problem easier by adding more variables.
  - Interested in samples from $P(x)$.
  - Introduce auxiliary variables $u_k$.
  - Form the joint distribution $P(x, u) = P(x)P(u|x)$.
  - Carefully select/design $P(u|x)$ so that:
    - $P(u|x)$ is easy to sample from;
    - $P(x|u)$ is also easy to sample from.

- Sample from the joint model by alternately sampling these conditionals.
- The marginals for $x$ match our desired distribution.
Generalised Swendsen-Wang

 Desired distribution: \( P(\mathbf{x}) = \frac{1}{Z_x} \pi_0(\mathbf{x}) \prod_k \Phi_k(\mathbf{x}) \).

 \( \pi_0(\mathbf{x}) \) is a simple base density, perhaps factorised over \( \{x_i\} \).

 Introduce \( k \) auxiliary variables \( u_k \), one for each \( \Phi_k \).

 Choose the following form for their conditional density,

 \[
P(\mathbf{u} | \mathbf{x}) = \prod_k \frac{1}{\Phi_k(\mathbf{x})} \mathcal{I} [0 \leq u_k \leq \Phi_k(\mathbf{x})]
\]  

(1)

 where \( \mathcal{I} \) is an indicator function which is 1 when its argument is true, and zero otherwise.

 Each \( u_k \) is independently and uniformly distributed from 0 to \( \Phi_k(\mathbf{x}) \).
Our choice of $P(u|x)$ leads to the following joint distribution,

$$P(x,u) = \frac{1}{Z_x} \pi_0(x) \prod_k I[0 \leq u_k \leq \Phi_k(x)]$$  \hspace{1cm} (2)$$

Note that the $\Phi_k$ terms from $P(x)$ and from $P(u|x)$ have canceled.

We also have

$$P(x|u) \propto \pi_0(x) \prod_k I[0 \leq u_k \leq \Phi_k(x)]$$  \hspace{1cm} (3)$$

I.e. $P(x|u)$ is just the ‘base’ density $\pi_0(x)$, restricted to the region satisfying the constraints $\{0 \leq u_k \leq \Phi_k(x)\}$. 
Potentials $\Phi_k(x_i, x_j)$ can only take two values.
$\Phi_k(x_i, x_j) = e^{W_{ij}}$ if $x_i = x_j$, and $\Phi_k(x_i, x_j) = e^{-W_{ij}}$ if $x_i = -x_j$.

Terms $\mathcal{I}[0 \leq u_k \leq \Phi_k(x)]$ may constrain the allowed combinations of $\{x_i\}$ when conditioning on $u$.

$W_{ij} > 0$
- If $u_k > e^{-W_{ij}}$ then we must have $x_i = x_j$.
- If $u_k \leq e^{-W_{ij}}$ then there is no direct constraint on $(x_i, x_j)$.

$W_{ij} < 0$
- If $u_k > e^{W_{ij}}$ then we must have $x_i = -x_j$.
- If $u_k \leq e^{W_{ij}}$ then there is no direct constraint on $(x_i, x_j)$.

Constraints give rise to “bound” clusters which act as a single unit.

We can replace $\{u_k\}$ with binary summary variables $b_k = \mathcal{I}[u_k > e^{-|W_{ij}|}]$ — denoting presence of a “bond”.
“Bond” Variables

❖ We have one binary bond variable, \( b_k \), for each weight \( W_{ij} \) in the original Boltzmann machine.

❖ Update bonds given \( \mathbf{x} \):
  ✦ If \( W_{ij} > 0 \), and if \( x_i = x_j \) then form bond \( (b_k = 1) \) with probability \( p = 1 - e^{-2W_{ij}} \), otherwise \( b_k = 0 \).
  ✦ If \( W_{ij} > 0 \), then ‘do the opposite’.

❖ If \( b_k = 1 \), then the units at the bond terminals must be:
  ✦ in the same states if \( W_{ij} > 0 \);
  ✦ in different states if \( W_{ij} < 0 \).

❖ Update \( \mathbf{x} \) given bonds:
  ✦ Identify connected cluster (units linked by active bonds).
  ✦ Each of these clusters has just two possible states. Select a state based on the probabilities that arise from the contribution of the residual \( \pi_0(\mathbf{x}) \) term — in a BM model this is the combined influence of the single unit biases.
Simple Example Of Update Procedure

- Each unit is connected to 4 nn with weight $W > 0$.
- Bonds form between units in the same state with probability $1 - e^{-2W_{ij}}$.
- After sampling the bonds, we have 6 connected clusters. Each cluster of units is independently resampled as a single block.
- New state differs considerably from the previous state.
Successive Realisations From An Ising Model

successive realizations

Gibbs

Swendsen - Wang

Figure 3. Successive Realizations of the Ising Model at Critical Temperature From the Single-Site Metropolis and Swendsen–Wang Algorithms.
References


❖ Higdon, D.M. *Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications* Journal of the American Statistical Association **93**(442), 1998. (Figures were cut from this paper.)

❖ There are also extended methods, for example “Partially Decoupled” SW (Higdon,93&98)